Edinburgh Research Explorer

Quantitative proteomics identifies tumour matrisome signatures in patients with non-small cell lung cancer

Download date: 13. Nov. 2023

Quantitative proteomics identifies tumour matrisome signatures in patients with non-small cell lung cancer

Helen F. Titmarsh[1,2], Alex von Kriegsheim[3], Jimi C. Wills[3], Richard A. O'Connor[2], Kevin Dhaliwal[2], Margaret C. Frame[3], Samuel B. Pattle[4], David A. Dorward[2,4], Adam Byron[3,5*†] and Ahsan R. Akram[2,3*†]

[1]The EPSRC and MRC Centre for Doctoral Training in Optical Medical Imaging, Queen's Medical Research Institute, University of Edinburgh, Edinburgh Bioquarter, Edinburgh, United Kingdom, [2]Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh Bioquarter, Edinburgh, United Kingdom, [3]Cancer Research UK Scotland Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom, [4]Department of Pathology, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom, [5]Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

Introduction: The composition and remodelling of the extracellular matrix (ECM) are important factors in the development and progression of cancers, and the ECM is implicated in promoting tumour growth and restricting anti-tumour therapies through multiple mechanisms. The characterisation of differences in ECM composition between normal and diseased tissues may aid in identifying novel diagnostic markers, prognostic indicators and therapeutic targets for drug development.

Methods: Using tissue from non-small cell lung cancer (NSCLC) patients undergoing curative intent surgery, we characterised quantitative tumour-specific ECM proteome signatures by mass spectrometry.

Results: We identified 161 matrisome proteins differentially regulated between tumour tissue and nearby non-malignant lung tissue, and we defined a collagen hydroxylation functional protein network that is enriched in the lung tumour microenvironment. We validated two novel putative extracellular markers of NSCLC, the collagen cross-linking enzyme peroxidasin and a disintegrin and metalloproteinase with thrombospondin motifs 16 (ADAMTS16), for discrimination of malignant and non-malignant lung tissue. These proteins were up-regulated in lung tumour samples, and high *PXDN* and *ADAMTS16* gene expression was associated with shorter survival of lung adenocarcinoma and squamous cell carcinoma patients, respectively.

Discussion: These data chart extensive remodelling of the lung extracellular niche and reveal tumour matrisome signatures in human NSCLC.

KEYWORDS

non-small cell lung cancer, matrisome, mass spectrometry, peroxidasin, ADAMTS16, lysine hydroxylation, lung, proteomic

Introduction

Lung cancer, including non-small cell lung cancer (NSCLC), is a common cancer and the leading cause of cancer-related deaths (1). Lung cancer, therefore, is a significant cause of morbidity and mortality, and improving lung cancer outcomes remains a clinically unmet need. Underpinning improved outcomes are a better understanding of the underlying biological process of tumour progression, identification of novel drug targets and establishment of markers that aid clinicians in determining diagnosis, prognosis and treatment decisions. Thus, proteomic assessment of primary cancer samples may help to advance these aims by uncovering differences in the protein composition of tumour and non-tumour tissues (2, 3).

The extracellular matrix (ECM) proteome, or matrisome, is composed of core ECM proteins, including collagens, glycoproteins and proteoglycans which form the structure of the ECM, and ECM-associated proteins, such as mucins, enzymes that can modify the ECM and secreted factors such as cytokines (4). Matrisome proteins have several key biological functions, including providing physical support, regulating pH, hydration and organisation of tissue and modulating signalling in tissue by binding growth factors and cytokines (5, 6). The matrisome is frequently altered in neoplastic tissues, impacting on multiple hallmark features of cancer (7, 8); for example, ECM dysregulation is associated with suppression of anti-tumour immunity and immunotherapy resistance, including in lung cancer (9, 10). Despite having important functions in cancer progression, the matrisome is an under-explored region of tumour tissues (11–13), owing in part to technical difficulties in its analysis, including the poor solubility of many ECM proteins (14), the multiplicity of their post-translational modifications (15), the limited number of robustly validated antibodies targeting ECM proteins (16) and low abundance of ECM proteins in many tissues compared to intracellular proteins (17). While bulk lung tumour tissue samples have been analysed by mass spectrometry (MS)-based proteomics to search for candidate biomarkers of disease (18–23), the tumour-associated matrisome of lung cancer patients has not been comprehensively documented. We hypothesised that deep, quantitative characterisation of ECM isolated from tumours from NSCLC patients will enable the detection of new prospective extracellular protein markers of lung cancer.

This study aimed to detect changes in matrisome proteins between human NSCLC tissues and patient-matched non-cancerous lung using a quantitative proteomic approach to curate an extensive resource of lung tumour matrisome proteins. We analysed tissue from 34 patients undergoing curative intent resections for NSCLC, coupling MS-based proteomics to fractionation of tissue samples to enrich for matrisome proteins. We quantified the differential abundance of proteins between tumour tissue and non-cancerous lung, characterising tumour matrisome signatures in NSCLC. Functional network analysis identified a collagen hydroxylation module enriched in tumour tissue, and we validated two novel putative extracellular markers, peroxidasin and a disintegrin and metalloproteinase with thrombospondin motifs 16 (ADAMTS16), as being up-regulated within tumours in separate patient cohorts.

Results

NSCLC proteomic study population characteristics

The cohort comprised 34 NSCLC patients, 20 female and 14 male, ranging from 49 to 87 years old, whom underwent surgical resections of NSCLC as treatment with curative intent (details in Supplementary Table 1). Seventeen of the patients were diagnosed with adenocarcinomas, 12 with squamous cell carcinomas, three with large cell tumours and two with pleomorphic tumours. Using TNM classification (8th edition), one patient had a T1, 17 had T2, 10 had T3 and six had T4 tumours (24). Lymph node (N) component included stage N0 in 24 patients, seven patients with N1 and three with N2 disease. As this was a curative intent cohort, no patients had distant metastasis. All patients except one were either previous or current cigarette smokers. The standardised uptake value of the PET imaging tracer 18F-FDG from the pre-operative CT scans of patients was high for all but two patients (one moderate, one not recorded). Non-cancerous lung tissue was retrieved from adjacent areas of resected lung tissue removed with the lung tumours (Figure 1A). Histopathological abnormalities within the non-malignant regions of the lung included single abnormalities or combinations of pneumonia, emphysema, inflammation, sarcoidosis and pleural fibrosis, and seven of the tissues were recorded as being histopathologically normal lung (Supplementary Table 1).

Quantification of NSCLC matrisome proteins by MS

To enrich extracellular proteins from patient-derived lung tissue for proteomic analysis, we used detergent and alkaline extractions and DNase treatment to deplete the lung tissue of cells and intracellular material, with minimal loss of relatively insoluble ECM proteins (Figure 1A; Supplementary Figure 1A). We performed label-free liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) analysis of ECM isolated from NSCLC tumour and non-tumour tissue, identifying a total of 1,662,859 peptide–spectrum matches, quantifying 3,602 proteins with a false discovery rate (FDR) of 1% (Supplementary Data 1). Extracellular proteins were significantly over-represented in the isolated ECM fractions (Figure 1B; Supplementary Data 2), and 1,420 identified proteins (39.4%) were annotated as extracellular in the Gene Ontology database (25) (Supplementary Datas 1, 2). The median sequence coverage of extracellular proteins, determined by MS, was significantly higher than that of all other identified proteins (Figure 1C), and the spectral counts for extracellular proteins were significantly higher than for other proteins (Figure 1D).

Two hundred and fifty-six proteins in the dataset were classified as curated matrisome proteins (26, 27), which comprised 108 core matrisome proteins and 148 matrisome-associated proteins. The core matrisome group consisted of 70 glycoproteins, 26 collagens and 12 proteoglycans; the matrisome-associated group consisted of 37 ECM-affiliated proteins, 73 ECM regulators and 38 secreted
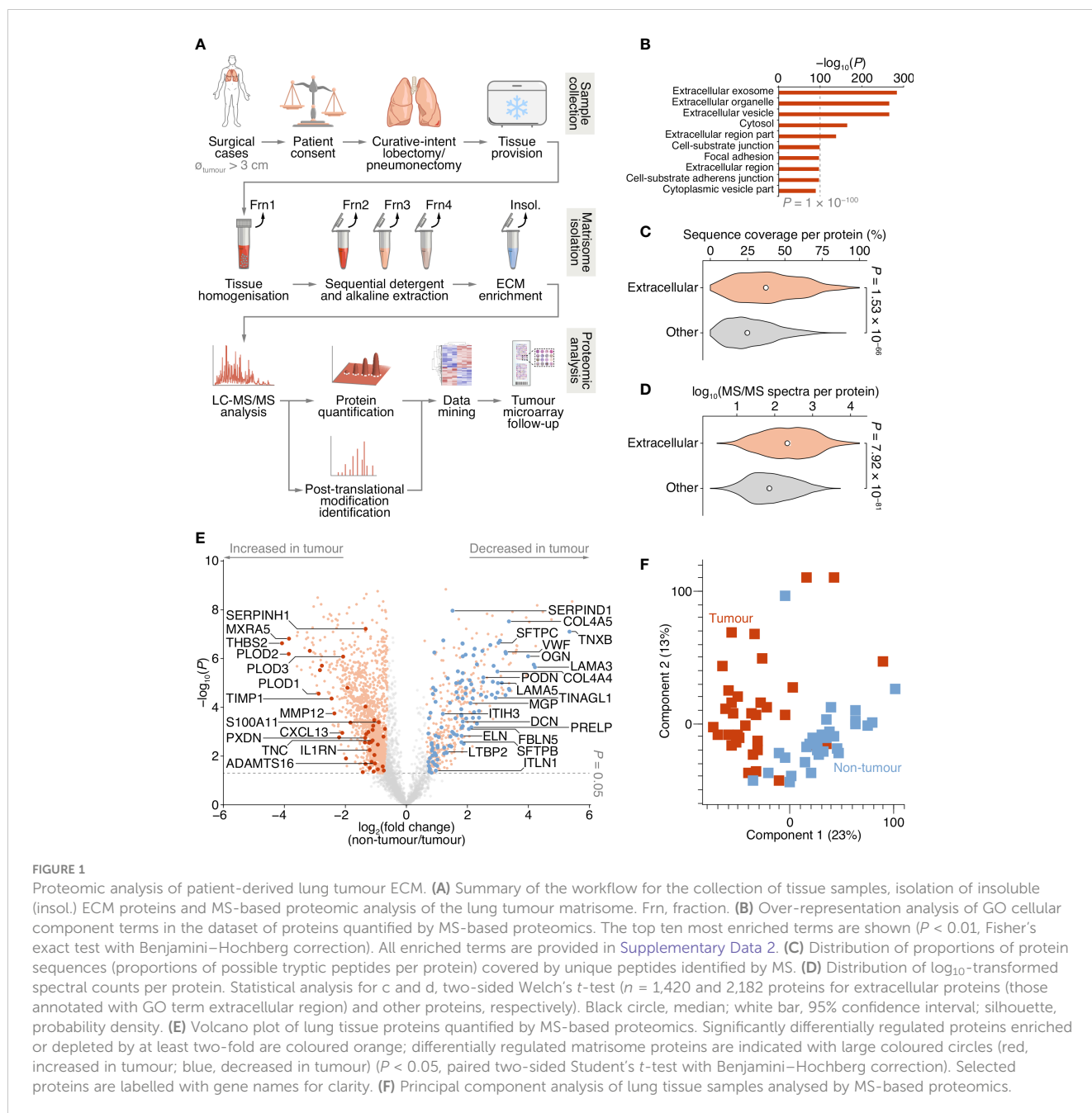
FIGURE 1

Proteomic analysis of patient-derived lung tumour ECM. **(A)** Summary of the workflow for the collection of tissue samples, isolation of insoluble (insol.) ECM proteins and MS-based proteomic analysis of the lung tumour matrisome. Frn, fraction. **(B)** Over-representation analysis of GO cellular component terms in the dataset of proteins quantified by MS-based proteomics. The top ten most enriched terms are shown ($P < 0.01$, Fisher's exact test with Benjamini−Hochberg correction). All enriched terms are provided in Supplementary Data 2. **(C)** Distribution of proportions of protein sequences (proportions of possible tryptic peptides per protein) covered by unique peptides identified by MS. **(D)** Distribution of $\log_{10}$-transformed spectral counts per protein. Statistical analysis for c and d, two-sided Welch's t-test ($n = 1,420$ and $2,182$ proteins for extracellular proteins (those annotated with GO term extracellular region) and other proteins, respectively). Black circle, median; white bar, 95% confidence interval; silhouette, probability density. **(E)** Volcano plot of lung tissue proteins quantified by MS-based proteomics. Significantly differentially regulated proteins enriched or depleted by at least two-fold are coloured orange; differentially regulated matrisome proteins are indicated with large coloured circles (red, increased in tumour; blue, decreased in tumour) ($P < 0.05$, paired two-sided Student's t-test with Benjamini−Hochberg correction). Selected proteins are labelled with gene names for clarity. **(F)** Principal component analysis of lung tissue samples analysed by MS-based proteomics.

proteins. This is comparable in number and matrisome subgroup distribution to recent proteomic analyses of lung tumour tissue (28–30) (Supplementary Figure 1B; Supplementary Table 2). Together, these data indicate strong detection and enrichment of extracellular proteins in the patient-derived lung ECM samples.

Of the 3,602 identified proteins, 1,805 significantly differed by at least 2-fold between tumour and non-tumour tissues ($P < 0.05$, paired two-sided Student's t-test with Benjamini–Hochberg correction) (Figure 1E; Supplementary Data 3). Pairwise correlation analysis of all tumour and non-tumour samples revealed positive correlation between like sample types (Supplementary Figure 1C), and tumour samples generally clustered together (Figure 1F; Supplementary Figure 1C). One hundred and sixty-one matrisome proteins were significantly

differentially expressed (Supplementary Data 4), 47 of which were increased in tumour samples compared to non-tumour samples, and 114 of which were decreased in tumour samples compared to non-tumour samples (Figure 2; Supplementary Figure 2).

Down-regulation of ECM-organising proteins in NSCLC

The majority of matrisome proteins up-regulated in tumour ECM were matrisome-associated proteins, predominantly ECM regulators and secreted proteins, whereas proteins down-regulated in tumour ECM included a substantial number of both matrisome-associated proteins and core matrisome proteins (Figure 2;
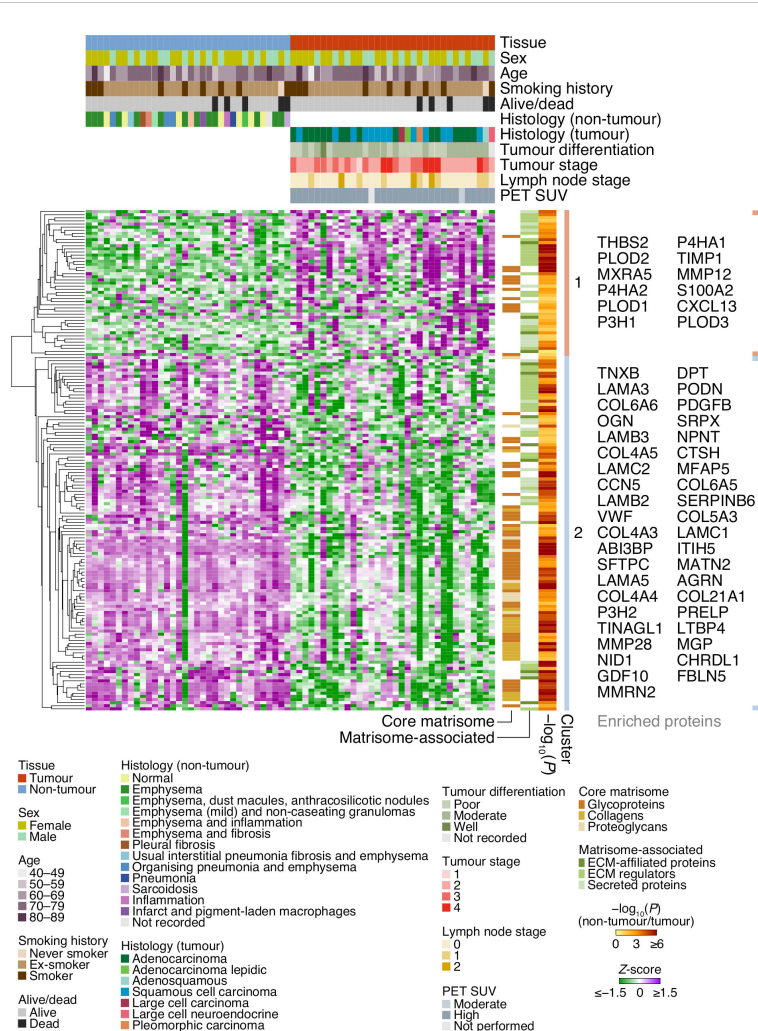
FIGURE 2

Cluster analysis of patient characteristics and matrisome protein expression. Heatmap representation of the 161 differentially expressed matrisome proteins ($P < 0.05$, paired two-sided Student's t-test with Benjamini−Hochberg correction). Proteins were quantified by $\log_2$-transformed label-free quantification intensities, standardised by row-wise Z-scoring and hierarchically clustered on the basis of Euclidean distance. The two principal clusters are indicated; proteins enriched or depleted by at least four-fold are labelled next to the respective cluster (selected proteins are labelled with gene names for clarity). SUV, standardised uptake value.

Supplementary Figure 2). Of those collagens that were differentially regulated, all were down-regulated in patient-derived tumour samples compared to matched non-tumour samples, ranging from 1.7-fold for type I collagen α1 chain (COL1A1) to 18.1-fold for type VI collagen α6 chain (COL6A6) (Supplementary Figure 2; Supplementary Data 3). In addition, all detected laminins were down-regulated in tumour samples compared to matched non-tumour samples, from 1.8-fold for laminin α2 chain (LAMA2) to 18.6-fold for laminin α3 chain (LAMA3) (Supplementary Figure 2; Supplementary Data 3). These findings agree with similar observations of decreased abundance of many collagen and laminin subunits in a murine model of lung adenocarcinoma (29).

Matrisome proteins down-regulated in tumours were enriched for ECM organisation functions (Supplementary Data 5), and these proteins formed a functional subnetwork dominated by collagens and laminins that clustered based on protein interactions (Figure 3A; Supplementary Figure 3A). Other down-regulated glycoproteins involved in ECM organisation clustered together,

including elastin (3.5-fold), EMILIN-1 (2.0-fold), fibulin-5 (4.1-fold) and microfibril-associated glycoprotein 4 (2.8-fold) (Figure 3A; Supplementary Figure 3A), all of which have been previously reported to be down-regulated in lung tumour ECM compared to normal ECM in mice (29). These data suggest that core matrisome proteins with key roles in ECM organisation and structural integrity are dysregulated in NSCLC tumours.

Increased lysine hydroxylation in fibrillar collagens in NSCLC

Matrisome proteins up-regulated in tumours were enriched for protein hydroxylation processes (Supplementary Data 5), driven by procollagen-lysine,2-oxoglutarate 5-dioxygenases (also known as lysyl hydroxylases) and prolyl 3- and prolyl 4-hydroxylases (Supplementary Figure 2; Supplementary Data 3), which formed a functional subnetwork of interacting proteins (Figure 3A;
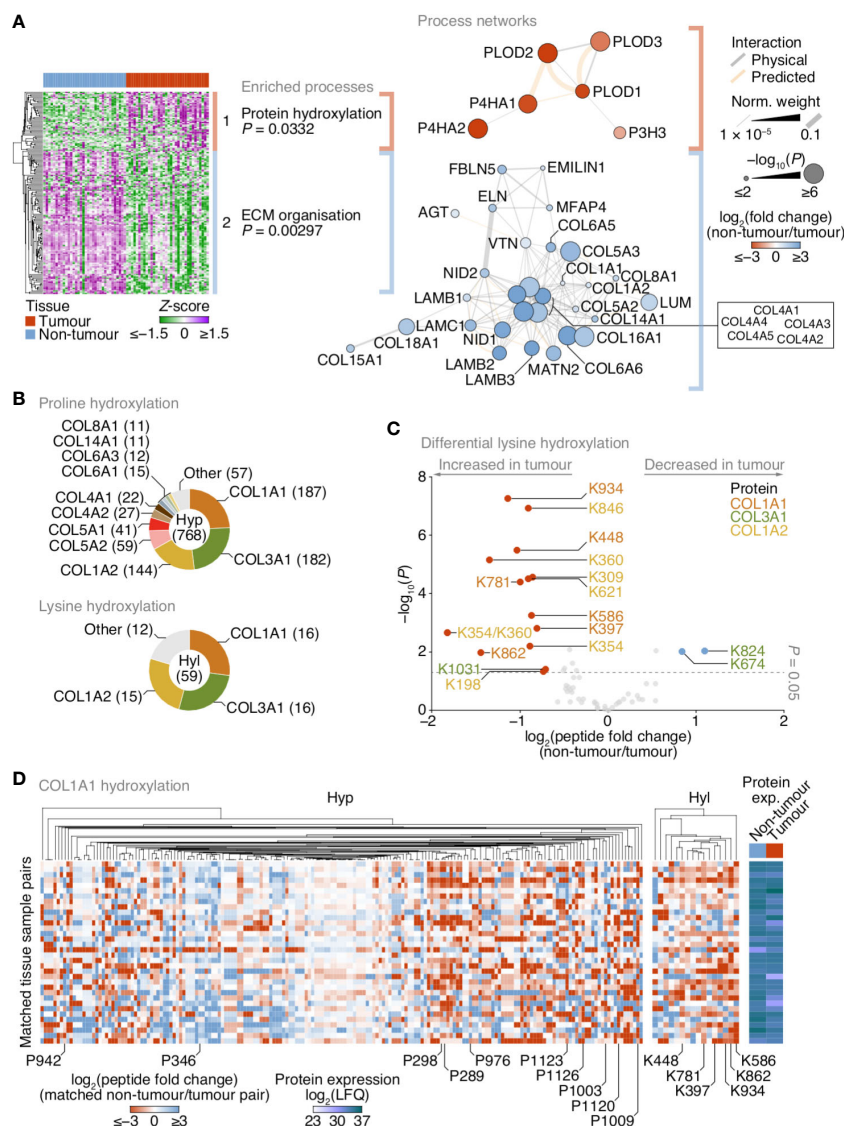
FIGURE 3

Up-regulation of lysine hydroxylation in patient-derived lung tumour ECM. **(A)** Gene ontology enrichment analysis of the principal clusters of differentially regulated matrisome proteins identified in Figure 2. All enriched terms for each respective cluster are shown (Fisher's exact test with Benjamini−Hochberg correction); for cluster 2, the term "extracellular structure organisation", the parent term of ECM organisation, and comprising the same contributing proteins and enriched with the same P-value, was omitted. Proteins belonging to enriched terms were used to construct corresponding protein interaction networks (right panels). Proteins (nodes) are coloured according to enrichment or depletion in tumour samples and sized according to statistical significance ($P < 0.05$, paired two-sided Student's t-test with Benjamini−Hochberg correction). Protein interactions (edges) were weighted according to evidence of co-functionality. Unconnected proteins are not shown. **(B)** Proportions of identified peptides containing hydroxylated proline (Hyp; top panel) or hydroxylated lysine (Hyl; bottom panel) assigned to corresponding proteins. Numbers of modified peptides are indicated in parentheses. **(C)** Volcano plot of peptides containing hydroxylated lysine quantified by MS-based proteomics. Differentially regulated peptides are indicated with large coloured circles (red, increased in tumour; blue, decreased in tumour) ($P < 0.05$, FDR 20%, paired two-sided Student's t-test with Benjamini−Hochberg correction). **(D)** Regulation of proline and lysine hydroxylation in type I collagen α1 chain (COL1A1) across 34 matched non-tumour−tumour paired tissue samples. The top ten most enriched peptides for each hydroxylation modification are labelled (all six enriched peptides containing hydroxylated lysine are labelled). Total protein expression determined by label-free quantification (LFQ) shown for corresponding samples. See Supplementary Data 6.

Supplementary Figure 3A). These proteins catalyse the post-translational formation of hydroxylysine and hydroxyproline residues in collagen chains, which are critical for collagen helix stability (31), and have been reported to be up-regulated in fibrotic lung ECM in murine and human model systems (32, 33). LC-MS/MS analysis revealed that the majority of hydroxylated residues were found in type I and type III collagen chains (COL1A1, COL1A2, COL3A1) (Figure 3B). While there were both up- and down-regulated hydroxyproline-containing peptides quantified by LC-MS/MS (Supplementary Figure 3B), almost all hydroxylysine-containing peptides were up-regulated in tumour ECM, and all-but-one of these were derived from type I collagen chains (Figure 3C).

To identify clusters of residues that showed similar regulation of hydroxylation across the tissue samples, we mapped the changes in abundance of hydroxyproline- and hydroxylysine-containing peptides for matched tumour−non-tumour pairs. This analysis

indicated variable modulation of proline hydroxylation in type I and type III collagen chains, including a substantial cluster of hydroxyproline-containing peptides derived from COL1A1 that were up-regulated in tumour ECM (Figure 3D; Supplementary Figures 3C, D). Lysine hydroxylation, however, was predominantly up-regulated in type I collagen chains across most matched tumour samples (Figures 3C, D; Supplementary Figure 3D), whereas modulation of lysine hydroxylation in type III collagen was more variable (Figure 3C; Supplementary Figure 3C). These data imply that type I collagen lysine residues are extensively hydroxylated in tumour ECM, in concert with an up-regulation of enzymes that catalyse collagen hydroxylation.

Up-regulation of matrisome-associated proteins in NSCLC

In addition to enrichment of enzymes that regulate ECM hydroxylation in tumours tissue, there was an increase in matrisome proteins associated with ECM turnover in tumour samples. Several cathepsins, some of which have reported roles in the degradation of extracellular proteins, were up-regulated in tumour samples, including the secreted thiol proteases cathepsin B (2.1-fold) and cathepsin S (1.7-fold) (Supplementary Figure 2; Supplementary Data 3). Several matrix metalloproteinases (MMPs) were up-regulated in tumour samples, namely MMP12 (5.1-fold), MMP14 (2.3-fold) and MMP2 (2.2-fold), which cleave elastin (and insulin), collagen (and other extracellular proteins, such as aggrecan) and certain collagens (and gelatin), respectively. In contrast MMP28, which cleaves casein, was down-regulated in tumour samples (7.3-fold) (Supplementary Figure 2; Supplementary Data 3). A disintegrin and metalloproteinase with thrombospondin motifs 16 (ADAMTS16), another protease, was up-regulated in tumour samples (2.0-fold).

In addition to many matrisome-associated proteins, several core matrisome glycoproteins were up-regulated in tumour tissue. The most enriched glycoproteins in tumour ECM compared to non-tumour ECM included thrombospondin-2 (16.9-fold), MXRA5 (14.4-fold), peroxidasin (2.5-fold), thrombospondin-1 (2.5-fold), SPARC (2.3-fold), FRAS1 (2.3-fold) and tenascin-C (2.1-fold) (Supplementary Figure 2; Supplementary Data 3). Four of the top five most up-regulated glycoproteins in tumour ECM – thrombospondin-1 and -2, SPARC and peroxidasin – formed a connected subnetwork (Supplementary Figure 3A), pointing to their common functional roles in organising the collagen ECM and regulating cell–ECM adhesion.

Peroxidasin and ADAMTS16 are increased in NSCLC tumour samples

Of the extracellular proteins we quantified as up-regulated in tumours, we selected peroxidasin and ADAMTS16 for further analysis (Figure 4A). The activity of peroxidasin – an enzyme that mediates collagen cross-linking – has recently been linked to the promotion of lung cancer through inhaled air-pollution particles,

which adsorb peroxidasin and induce aberrant ECM thickening (34), providing a rationale to examine peroxidasin expression in early-stage NSCLC ECM. *ADAMTS16* DNA methylation has been shown to be dysregulated in several epithelial cancers, including lung cancer (35), suggesting that regulation of ADAMTS16 protein expression in tumour-associated ECM, and its potential as a protein biomarker in NSCLC, warrant investigation. We constructed a tumour microarray (TMA) containing control and tumour samples from a separate cohort of lung cancer patients undergoing curative intent surgery (Supplementary Table 3). Selected proteins were analysed by immunohistochemistry (IHC) and scored on a four-point scale (scored from 0–3) based on increasing immunostaining by IHC from no staining to strong staining. For lung cores in the TMA derived from tumour samples, the distribution and scoring of immunostained proteins was assessed separately for tumour stroma and tumour cells. In tumours, peroxidasin and ADAMTS16 were found variably distributed between tumour stroma and tumour cells (Figure 4B). Importantly, both proteins had significantly increased immunostaining intensities in tumour cells compared to non-tumour cells (Figures 4B, C). These data confirm the results from the MS-based proteomics analysis and imply that peroxidasin and ADAMTS16 are up-regulated in NSCLC tumours.

To examine the expression of the genes encoding the selected extracellular proteins, we analysed an integrated dataset of RNA-seq data derived from paired tumour and adjacent normal tissue from lung adenocarcinoma or lung squamous cell carcinoma patients (36). Expression of both *PXDN* (which encodes peroxidasin) and *ADAMTS16* were significantly up-regulated in adenocarcinoma tumours compared to matched adjacent normal lung tissue (1.96-fold and 10.7-fold change in median expression, respectively) (Figure 4D). Both genes were also significantly up-regulated in squamous cell carcinoma tumours compared to matched adjacent normal lung tissue (1.78-fold and 5.62-fold change in median expression, respectively) (Figure 4E). These data indicate that *PXDN* and *ADAMTS16* are transcriptionally up-regulated in NSCLC tumours.

We next assessed whether the protein expression levels of peroxidasin and ADAMTS16 were predictors of survival in the TMA cohort (Supplementary Table 3). Although up-regulated in tumour tissue (Figures 4A–C), the degree of tumour stroma immunostaining of these two markers, as determined by IHC scoring of all interpretable lung cores, did not significantly correlate with survival outcome within the limited TMA cohort (Supplementary Figure 4). To examine patient survival in larger NSCLC cohorts, we used the Cancer Genome Atlas (TCGA) data derived from primary tissue samples from lung adenocarcinoma (513 patients) or lung squamous cell carcinoma (501 patients). We found that higher expression of *PXDN* was significantly associated with shorter survival of adenocarcinoma patients (hazard ratio (HR) = 1.57; 95% confidence interval (CI), 1.05–2.37; $P = 0.028$, log-rank test), whereas that of *ADAMTS16* was not (Figure 4F). Median survival of adenocarcinoma patients was 54.4 months for the low *PXDN* expression subgroup and 39.0 months for the high *PXDN* expression subgroup. In contrast, there was no statistical association of expression of *PXDN* with survival of squamous cell carcinoma
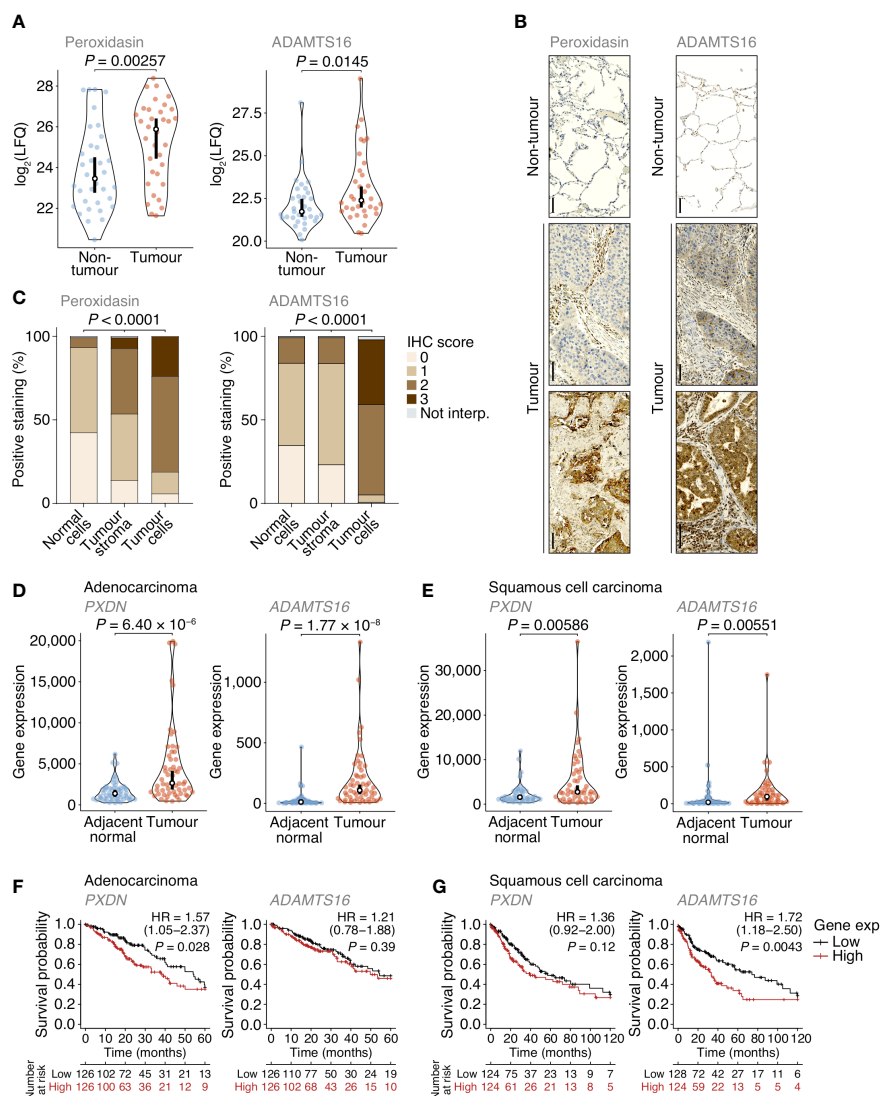
FIGURE 4

Dysregulation of peroxidasin and ADAMTS16 in NSCLC tumour cells. **(A)** Label-free quantification of peroxidasin (left panel) and ADAMTS16 (right panel) in non-tumour and tumour ECM samples quantified by MS-based proteomics. Statistical analysis, paired two-sided Student's t-test with Benjamini−Hochberg correction ($n = 34$ matched non-tumour−tumour paired tissue samples). **(B)** Immunohistochemistry (IHC) anaysis of peroxidasin (left panels) and ADAMTS16 (right panels) for exemplar lung cores from the tumour microarray. Scale bars, 100 µm. **(C)** Quantification of TMA immunostaining for non-tumour-derived lung cores (normal cells) and tumour-derived lung cores (tumour stroma and tumour cells). IHC scores of 0, 1, 2 and 3 correspond to no, mild, moderate and strong immunostaining, respectively. Not interp., not interpretable. Statistical analysis, chi-square test. **(D, E)** Differential expression analysis of *PXDN* (left panels) and *ADAMTS16* (right panels) in cohorts of lung adenocarcinoma **(D)** or lung squamous cell carcinoma **(E)** patients from integrated RNA-seq datasets. Statistical analysis, Wilcoxon signed-rank test ($n = 57$ and 49 matched adjacent normal−tumour paired tissue samples for adenocarcinoma and squamous cell carcinoma patients, respectively). For (A, D, E) black circle, median; black bar, 95% confidence interval; silhouette, probability density. **(F, G)** Kaplan−Meier curves of patient survival associated with degree of expression of *PXDN* (left panels) and *ADAMTS16* (right panels) in cohorts of lung adenocarcinoma **(F)** or lung squamous cell carcinoma **(G)** patients from TCGA RNA-seq datasets. The lower quartiles (low) and upper quartiles (high) of gene expression were compared. HR, hazard ratio (with 95% confidence interval). Statistical analysis, log-rank test ($n = 513$ and 501 patients for adenocarcinoma and squamous cell carcinoma cohorts, respectively).

patients, whereas higher expression of *ADAMTS16* was significantly associated with shorter survival of squamous cell carcinoma patients (HR = 1.72; 95% CI, 1.18–2.50; $P = 0.0043$) ([Figure 4G](#)). Median survival of squamous cell carcinoma patients was 74.1 months for the low *ADAMTS16* expression subgroup and 33.4 months for the high *ADAMTS16* expression subgroup. Together, these results indicate that these proteins, identified in the patient lung tumour matrisome, are increased in NSCLC and their corresponding gene expression provides prognostic information for lung adenocarcinoma (*PXDN*)

or squamous cell carcinoma (*ADAMTS16*) patients undergoing curative intent surgery.

Discussion

We present herein a comprehensive characterisation of the NSCLC matrisome. To our knowledge, this is the first unbiased matrisome-scale proteomic analysis of enriched ECM from lung

cancer patient tissue. Our quantitative analyses reveal the differential expression of a substantial number of extracellular proteins in lung tumour tissue from patients with early-stage NSCLC, implying extensive remodelling of the extracellular niche.

We found that many core matrisome proteins were less abundant in lung tumour tissue as compared to non-cancerous lung tissue. For example, all differentially expressed collagens and proteoglycans were relatively decreased in tumour samples, while a subset were not significantly altered. These findings are consistent with MS-based proteomic data from a murine model of lung adenocarcinoma, which identified decreased or little change in abundance of many core matrisome proteins, including the majority of detected collagens and laminins, in tumour ECM as compared to non-tumour ECM (29).

The general association of lung cancer with desmoplasia, however, suggests that our observations may be linked to early stages of tumour ECM remodelling prior to accumulation of collagen. Indeed, recent examination of idiopathic pulmonary fibrosis lung tissue identified induction of collagen-modifying enzymes that contribute to collagen cross-linking, including lysyl hydroxylase 2, but not increased collagen synthesis, as a defining feature of lung fibrosis that increases tissue stiffness and promotes fibrotic progression (37, 38). Lysyl hydroxylase 2 (encoded by *PLOD2*) is one of several collagen-modifying enzymes that we detected as up-regulated in NSCLC tumour tissue in this study, alongside a concomitant increase in collagen lysine hydroxylation. Lysyl hydroxylase 2 is secreted by lung cancer cells in culture, and its hydroxylation of collagen telopeptidyl lysine residues leads to the formation of stable hydroxylysine aldehyde-derived collagen cross-links that are up-regulated in lung cancer tissue and generate stiffer tumour tissue (39, 40). Thus, dysregulation of collagen architecture or cross-linking, and consequential ECM stiffness, may be more prominent features of early-stage NSCLC than changes in total collagen synthesis or density.

We observed the up-regulation of enzymes involved in extracellular protein degradation in lung tumour tissue, including several cathepsins and MMPs. These proteinases target a wide range of ECM proteins, such as collagen, laminin and elastin (41, 42), almost all of which were depleted in tumour samples. This suggests that there may be increased turnover of core ECM macromolecules in NSCLC tumours, consistent with the remodelling of the extracellular niche found in various respiratory diseases and invasive tumour growth (41, 43). Indeed, the expression of MMP2 and MMP14, which were up-regulated in lung tumour ECM, is linked to poorer patient outcomes in NSCLC (44, 45), and MMP12, also up-regulated in tumour samples, is associated with faster disease relapse and metastasis in NSCLC patients (46) and the occurrence of bronchioalevolar adenocarcinomas in patients with emphysema (47).

We validated in a separate patient cohort the up-regulation in tumour tissue of two extracellular proteins identified by MS-based proteomics. Peroxidasin, an extracellular peroxidase, mediates the formation of sulfilimine cross-links between methionine and hydroxylysine residues in type IV collagen (48, 49). Interestingly, the activity of peroxidasin is associated with accelerated tumorigenesis in murine models of lung carcinoma in the

presence of inhalable fine particulate matter, which adsorbs peroxidasin and leads to accumulation of collagen cross-linking (34). In our analyses, peroxidasin clustered in a tumour-enriched collagen-modifying protein subnetwork, which together with our identification of a tumour-enriched collagen hydroxylation functional network, implies that collagen modifications and modulation of collagen cross-linking are key characteristics of early-stage NSCLC. The other selected extracellular tumour marker candidate, ADAMTS16, a matrisome-associated protease, targets fibronectin to inhibit ECM assembly (50), further suggesting that ECM organisation is dysregulated in NSCLC tissue. IHC analyses determined that both peroxidasin and ADAMTS16 were enriched in tumour cell-rich regions of lung tumour tissue, although IHC scoring of either of these candidates did not represent a significant predictor of survival in the TMA cohort of NSCLC patients. The assessment of patients with early-stage disease undergoing curative intent surgery in this study precludes the detection of matrisome protein changes that occur in more advanced disease, which could limit the ability to identify robust late-stage disease biomarkers but likely also explains the identification of putative early events in the remodelling of collagen cross-linking. In addition, tissue was only collected from a small portion of the tumour. Tumours are known to be heterogeneous (51, 52); therefore, sampling from one area may not be representative of all the pathological protein changes within a tumour. Analysis of transcriptomic data from larger lung cancer patient cohorts, however, revealed increased tumour expression of genes encoding both peroxidasin and ADAMTS16, and this was separately linked to poorer survival of lung adenocarcinoma and squamous cell carcinoma patients, respectively. Recent analysis of matrisome gene expression in multiple transcriptomic datasets showed that extracellular protein levels are generally concordant with corresponding gene expression in human tissues (53). Our findings are consistent with this observation and suggest that high expression of *PXDN* and *ADAMTS16* can be used as surrogate readouts for up-regulation of these two potential extracellular lung tumour markers.

In summary, this study provides an extensive analysis of lung tissue ECM in patients with lung cancer, charting the remodelling of the matrisome in early-stage NSCLC. We show that proteomic profiling of patient-derived lung tumour ECM enables the identification of candidate extracellular markers of tumour cells. In addition, the systems-level changes to the lung matrisome we report here, including the up-regulation of a collagen hydroxylation network in NSCLC tissue, reveal potential molecular networks that could modulate ECM organisation and regulate lung cancer progression.

Materials and methods

Study approvals

Patient samples were collected with ethical approval and written patient consent. Ethical approval was granted by Lothian NRS Bioresource, REC number 15/ES/0094 (reference SR419). All

samples were assigned an anonymised code, and researchers were blinded to patient details for experiments. The TMA dataset was approved by Lothian NRS Bioresource, REC number 15/ES/0094 (reference SR1208), and approved by the NHS Lothian Caldicott Guardian (reference CRD19031).

Patient samples

Tissues samples used for mass spectrometry (MS) were collected from patients with NSCLC undergoing curative-intent surgical procedures. Following resection, samples were handled by an experienced thoracic pathologist, and samples of tumour and non-cancerous lung (from the most distal portion of the resection specimen) were dissected and provided for MS analysis. Samples were snap frozen and stored at $-70°C$ until required. Anonymised patient details were recorded, including age, gender, smoking history, histopathological diagnosis of tumour and non-tumour tissues, degree of differentiation of the tumour tissue, tumour stage and lymph node stage (determined by TNM status), survival and PET tracer uptake.

Enrichment of matrisome proteins

Tissue samples were enriched for predominately insoluble matrisome proteins by depleting soluble intracellular proteins (Figure 1A). Methods were adapted from previously published work (54). Tissue samples were finely minced with scalpels, and the presence of any necrosis and tissue pigmentation was noted. Samples were homogenised twice in 1 ml of chilled phosphate-buffered saline (PBS) (without Ca^{2+} or Mg^{2+}), containing 1% (v/v) protease inhibitor cocktail (Sigma-Aldrich), using a Precellys 24 tissue homogeniser (Bertin Instruments) at 6,500 rpm for 50 s. Samples were incubated for 5 min on ice between homogenisation cycles. The homogenate was centrifuged at $14,000 \times g$ for 10 min at 4°C. The supernatant was removed (fraction 1), and the remaining pellet was resuspended and incubated in 10 mM Tris-HCl, pH 8, 150 mM NaCl, 25 mM EDTA, 1% (v/v) Triton X-100, 1% (v/v) protease inhibitor cocktail for 30 min on ice. The lysate was centrifuged at $14,000 \times g$ for 10 min at 4°C. The supernatant was removed (fraction 2), and the remaining pellet was resuspended and incubated in 20 mM NH_4OH containing 0.5% (v/v) Triton X-100 in PBS (without Ca^{2+} or Mg^{2+}). The lysate was centrifuged at $14,000 \times g$ for 10 min at 4°C. The supernatant was removed (fraction 3), and the remaining pellet was incubated with 10 μg/ml DNase I in PBS (with Ca^{2+} and Mg^{2+}) for 30 min on ice. Samples were centrifuged at $14,000 \times g$ for 10 min at 4°C, and the supernatant was removed (fraction 4). The remaining pellet was washed in ice-cold PBS and centrifuged at $14,000 \times g$ for 10 min at 4°C three times. The final insoluble pellets enriched for ECM proteins were then stored at $-70°C$ until further use.

SDS-PAGE and western blotting

To confirm matrisome proteins were enriched in the final protein pellet prior to proteomics experiments, the supernatants from serially extracted fractions from a subset of samples were probed by SDS-PAGE and western blotting. Three percent of the supernatant from each fraction was used for western blotting to allow comparison between fractions. Fractions 1–4 were incubated in 1× Laemmli buffer containing 50 mM dithiothereitol for 10 min at 95°C. The final ECM pellet was precipitated using TCA–acetone (see below), homogenised and incubated in 8 M urea, 100 mM NH_4HCO_3, pH 8, 10 mM dithiothereitol for 30 min at 37°C. Samples were resolved by SDS-PAGE using 4–20% Tris-glycine gels. Proteins were transferred to PVDF membranes using an iBlot 2 dry blotting system (Thermo Fisher Scientific) according to manufacturer's instructions. Membranes were blocked using milk blocking buffer (5% (w/v) non-fat skimmed milk powder (Marvel) in 1× Tris-buffered saline containing 0.1% (v/v) Tween 20 (TBS-T)) for 1 h at room temperature with shaking. Following blocking, membranes were incubated with primary antibodies, diluted 1:1,000 in milk blocking buffer, overnight at 4°C with rolling. Antibodies used were rabbit polyclonal anti-fibronectin (#ab2413, Abcam), mouse monoclonal anti-lamin A/C (clone 4C11; #4777, Cell Signaling Technology) and rabbit monoclonal anti-GAPDH (clone 14C10; #2118, Cell Signaling Technology). Membranes were washed three times in TBS-T and then incubated with anti-rabbit or anti-mouse horseradish peroxidase-conjugated secondary antibodies, diluted 1:10,000 or 1:5,000, respectively, in milk blocking buffer, for 45 min at room temperature with rolling. Membranes were washed three times in TBS-T. Membranes was developed using Pierce ECL western blotting substrate (Thermo Fisher Scientific) according to the manufacturer's instructions. Membranes were imaged using an Odyssey Fc imaging system (LI-COR Biosciences).

Matrisome protein precipitation

Ice-cold TCA (10% (v/v) final concentration) was added to final ECM-enriched fractions for 20 min at 4°C. The sample was centrifuged at $16,000 \times g$ for 30 min at 4°C. The supernatant was discarded, and the ECM-enriched pellet was resuspended in ice-cold acetone, using vortexing and sonication, and incubated for 20 min at $-20°C$. The sample was centrifuged at $16,000 \times g$ for 30 min at 4°C, and the supernatant was discarded. The acetone wash step was then repeated. Samples were air dried at room temperature until residual solvent had evaporated. Precipitated protein pellets were stored at $-70°C$ until further use.

Matrisome protein digestion

Precipitated protein pellets were resuspended in 300 μl solubilisation buffer (8 M urea, 100 mM NH_4HCO_3, pH 8, 10 mM dithiothreitol). Samples were sonicated on ice using a probe sonicator and then incubated for 30 min at 37°C. After cooling to room temperature, sample pH in the range pH 8–9 was verified using pH indicator paper. To 50 μl sample, 8.3 μl of 175 mM iodoacetamide in 100 mM NH_4HCO_3, pH 8, was added (final concentration 25 mM) and incubated for 30 min at room

temperature in the dark to block thiol groups of cysteine residues. Excess iodoacetamide was quenched by the addition of 3 µl of 100 mM dithiothreitol in 100 mM NH_4HCO_3, pH 8. For protein digestion, urea was diluted from 8 M to 2 M using 100 mM NH_4HCO_3, pH 8. Samples were incubated with 833 units of PNGase F for 2 h at 37°C with shaking (900 rpm), then with 800 ng of Lys-C for 2 h at 37°C, then with 1.6 µg of MS-grade trypsin for 16 h at 37°C with shaking (1,200 rpm). Samples were then incubated with an additional 0.8 µg of trypsin for 2 h at 37°C with shaking (1,200 rpm). Samples were acidified using trifluoroacetic acid (TFA) to obtain sample pH in the range pH 3–4, and samples were clarified by centrifugation at 18,000 × g for 15 min.

Peptide desalting prior to MS analysis

Peptide concentrations in the digested samples were estimated using a Nanodrop spectrophotometer measuring absorbance at 280 nm. Stop-and-go extraction (Stage) tips were made in-house, using a method adapted from Rappsilber et al. (55). Briefly, using an 18-gauge blunt-ended needle, two disks were cut from C18 solid-phase extraction material and placed on top of each other inside a 200-µl EasyLoad pipette tip (Greiner Bio-One). Stage tips were loaded into a custom-built tip holder over a deep 96-well plate. Methanol was added to each Stage tip, and tips were centrifuged at 300 × g for 2 min. Stage tips were then equilibrated with 0.1% (v/v) TFA and centrifuged at 500 × g for 5 min. Samples estimated to contain 10 µg of acidified peptide (based on Nanodrop readings) were then added to Stage tips, and tips were centrifuged at 500 × g for 5 min. Stage tips with bound protein were stored at −20°C until elution.

Desalted peptide elution

Peptides were eluted from C18-containing Stage tips with 40 µl of 80% (v/v) acetonitrile, 0.1% (v/v) TFA, and tips were centrifuged at 200 × g for 5 min. Acetonitrile was evaporated using a vacuum centrifuge. Samples were adjusted to 15 µl volume with 0.1% (v/v) TFA, and peptide concentration was re-measured using a Nanodrop spectrophotometer measuring absorbance at 280 nm.

MS data acquisition

'Bottom-up' liquid chromatography-coupled tandem MS (LC-MS/MS) was used to elucidate the structure of isolated peptides and to detect post-translational modifications. LC-MS/MS analysis was carried out using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to an UltiMate 3000 UHPLC Nano (Thermo Fisher Scientific), Aurora C18 column (IonOpticks), column oven (maintained at 50°C; Sonation) and Proxeon Nanospray ion source (Thermo Fisher Scientific). Peptides (1 µg) were injected onto an Aurora C18 column in buffer A (2% (v/v) acetonitrile, 0.5% (v/v) acetic acid) and eluted with a linear 120-min gradient of 2%–45% (v/v) buffer B (80% (v/v) acetonitrile, 0.5%

(v/v) acetic acid). Eluting peptides were ionised in positive-ion mode before data-dependent analysis. A dynamic exclusion window of 30 s was enabled and lockmass was not used.

MS data analysis

The MS data were normalised and quantified using MaxQuant software (version 1.6.10.43) (56). MaxQuant quantifies proteins using a label-free technique, which calculates a normalised peptide abundance from ion signal intensities (57). Peptide lists were searched against the human UniProt knowledgebase database (version 2019_09) (58) and a common contaminants database using the Andromeda search engine implemented in MaxQuant. Cysteine carbamidomethylation was set as a fixed modification, and methionine oxidation, lysine oxidation, proline oxidation, N-terminal deamidation and protein N-terminal acetylation were set as variable modifications, with up to five modifications per peptide. Peptide identifications were matched between runs if they eluted within a time window of 0.7 min. Enzyme specificity was C-terminal to arginine and lysine, except when followed by proline. A maximum of two missed cleavages were permitted in the database search; minimum peptide length was seven amino acids. At least two peptide ratios were required for label-free quantification, and large label-free quantification ratios were stabilised. Peptide and protein false-discovery rates (FDRs) were set to 1%, determined by applying the target-decoy search strategy implemented in MaxQuant. Proteins matching to the common contaminants or decoy databases, and matches only identified by site, were omitted. Label-free quantification intensities were $\log_2$ transformed, and proteins quantified in less than one-third of samples were removed. Missing values were imputed from a width-compressed, down-shifted Gaussian distribution using Perseus (version 1.6.2.3) (59).

Functional enrichment analysis

Gene Ontology (GO) over-representation analysis was performed using DAVID 2021 (DAVID Knowledgebase, version 2022q2) (60). The functional annotation tool category of GO term enrichment was used to filter out very broad GO terms. GO over-representation analysis of matrisome proteins used the entire matrisome database as background. Significant over-representation of terms was determined using a Fisher's exact test with Benjamini–Hochberg correction.

Matrisome data analysis

Identified proteins were defined as belonging to the matrisome if they exist in searchable databases of matrisome proteins based on data from 17 studies of the ECM, MatrisomeDB (27), and *in-silico* and *in-vivo* data from the Matrisome Project (26). Label-free quantification intensities for proteins derived from tumour and non-tumour samples were compared using a paired two-sided Student's *t*-test, with FDR set to 5% and artificial within-groups

variance (s0) set to 1 using Perseus. Statistical data were visualised using Prism (version 9.2.0; GraphPad). For differentially expressed matrisome proteins, potential confounding variables were also considered. These variables included necrotic tumour samples versus non-necrotic tumour samples, adenocarcinoma versus squamous cell carcinoma, well and moderately differentiated tumours versus poorly differentiated tumours, tumours with local lymph node spread versus no known lymph node metastasis and normal non-tumour samples versus non-tumour samples with underlying pathology. Necrosis was crudely assessed based on the composition of the tumour samples when dissected and minced. PET standardised uptake value was scored as mild, moderate or marked. In some instances, absolute values were provided in reports rather than categorical values. These numbers were reclassified as no uptake for 0.6–0.8, low uptake for 1.0–2.0, moderate uptake for 1.5–2.0 and high uptake for >2.5. These variables were compared using multiple Student's t-tests ($P < 0.05$, FDR 5%).

Hierarchical cluster analysis

To enable relative comparison of protein enrichment, $\log_2$-transformed label-free quantification intensities were standardised by row-wise (protein-wise) Z-scoring. Differentially expressed matrisome proteins were hierarchically clustered on the basis of Euclidean distance, computed using average linkage, using Cluster 3.0 (C Clustering Library, version 1.54) (61). The following variables were included in the protein enrichment analysis: sex, age, smoking history, if patients had died at the time of analysis, histopathology results, degree of tumour differentiation, tumour stage, lymph node stage and PET 18F-FDG tracer uptake. Modified peptide fold changes between matched non-tumour–tumour pairs were hierarchically clustered on the basis of Euclidean distance, computed using average linkage. For sample correlation analysis, Spearman rank correlation coefficient-based distance matrices were computed using average linkage. Clustering results were visualised using Java TreeView (version 1.1.5r2) (62).

Interaction network analysis

Composite functional association networks were constructed using GeneMANIA (version 3.5.2; human interactions) (63) in Cytoscape (version 3.8.0) (64). Networks were based on reported physical and predicted protein–protein interactions; edges, representing protein–protein interactions, were weighted according to evidence of co-functionality using GeneMANIA. Networks were clustered using Markov clustering (granularity parameter 2.5), and graph layouts were determined using the force-directed algorithm in the Prefuse toolkit (65).

TMA and immunohistochemistry

The TMA was constructed for sequential patients undergoing surgical resection for lung cancer over a 2-year period at a regional thoracic surgery centre. Formalin-fixed paraffin pathological blocks were annotated by an experienced pathologist, and 1-mm cores taken from tumours and non-cancerous lung were embedded into new blocks and, subsequently, 4-μm sections were cut onto glass slides. Slides were deparaffinised and rehydrated, and antigen retrieval was undertaken with citrate buffer (#ab64214, Abcam) twice for 5 min in a microwave. Slides were processed with a commercial DAB cell and tissue staining kit (#CTS019, R&D Systems). Primary antibody immunostaining was optimised for polyclonal rabbit anti-human peroxidasin (#abx101905, Abbexa) and polyclonal rabbit anti-human ADAMTS16 (#TA322059, AMS Biotechnology) (diluted 1:100 and 1:200, respectively, and incubated at room temperature for 1 h). Secondary antibodies were incubated at room temperature for 30 min and DAB was developed. Slides were counterstained with haematoxylin and mounted, and images were acquired on an Axioscan microscope slide scanner (Zeiss).

For peroxidasin immunostaining, there were 151 non-cancerous lung cores that were available for evaluation (149 interpretable following staining) and 138 tumour cores, of which 119 were paired samples (Supplementary Table 3). For ADAMTS16 immunostaining, there were 150 available non-cancerous lung cores (148 interpretable for inflammatory cells and 149 for non-tumour cells) and 155 available tumour cores (154 interpretable for tumour stroma and 152 for tumour cells), of which 133 were paired samples (Supplementary Table 3). Slides were scored for intensity of staining of tumour cells or stromal areas on a four-point scale: 0, no staining; 1, low staining; 2, moderate staining; 3, high staining. Data are presented as the proportion of positive staining within each category of staining for tumour and normal samples. Significant difference in the distribution of staining scores was determined by a chi-square test. Outcome data were recorded for each patient who had a tumour resection, with the median time to follow-up of 1,432 days (range 1,054–1,905 days). Outcome included a record of death as defined by the clinical care team. Immunohistochemistry (IHC) scoring was assessed against survival for tumour staining.

Differential gene expression analysis

Gene expression data were derived from RNA-seq datasets of primary tissue samples from cohorts of lung adenocarcinoma or lung squamous cell carcinoma patients extracted from TCGA, the Genotype-Tissue Expression repository and the Therapeutically Applicable Research to Generate Effective Treatments database using TNMplot (36). Data from tumour samples were compared to paired data from matched adjacent normal samples.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD041066.

Survival analysis

Kaplan–Meier curves were computed from RNA-seq datasets of primary tissue samples from cohorts of lung adenocarcinoma or lung squamous cell carcinoma patients extracted from TCGA using KMplot (66). The lower and upper quartiles of gene expression were compared. Follow-up time was truncated to retain at least 10 patients at risk.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://proteomecentral. proteomexchange.org, PXD041066.

Ethics statement

Patient samples were collected with ethical approval and written patient consent. Ethical approval was granted by Lothian NRS Bioresource, REC number 15/ES/0094 (reference SR419). All samples were assigned an anonymised code, and researchers were blinded to patient details for experiments. The TMA dataset was approved by Lothian NRS Bioresource, REC number 15/ES/0094 (reference SR1208), and approved by the NHS Lothian Caldicott Guardian (reference CRD19031).

Author contributions

HT, AK, JW, RO'C, AB and AA undertook experimentation and data generation. HT, AK, JW, RO'C, KD, MF, AB and AA interpretated the data and conceptualised the manuscript. AA, SP, DD constructed the TMA, retrieved the clinical data and provided samples for ECM extraction. HT, AB and AA wrote the manuscript. AB and AA jointly supervised the work. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1194515/ full#supplementary-material

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/ caac.21492

2. Simpson RJ, Bernhard OK, Greening DW, Moritz RL. Proteomics-driven cancer biomarker discovery: looking to the future. *Curr Opin Chem Biol* (2008) 12:72–7. doi: 10.1016/j.cbpa.2008.02.010

3. Indovina P, Marcelli E, Pentimalli F, Tanganelli P, Tarro G, Giordano A. Mass spectrometry-based proteomics: the road to lung cancer biomarker discovery. *Mass Spectrom Rev* (2013) 32:129–42. doi: 10.1002/mas.21355

4. Hynes RO, Naba A. Overview of the matrisome–an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* (2012) 4:a004903. doi: 10.1101/cshperspect.a004903

5. Hynes RO. The extracellular matrix: not just pretty fibrils. *Science* (2009) 326:1216–9. doi: 10.1126/science.1176009

6. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. *J Cell Sci* (2010) 123:4195–200. doi: 10.1242/jcs.023820

7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013

8. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep* (2014) 15:1243–53. doi: 10.15252/embr.201439246

9. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF-β- associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nat Commun* (2018) 9:4692. doi: 10.1038/s41467- 018-06654-8

10. Peng DH, Rodriguez BL, Diao L, Chen L, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8+ T cell exhaustion. *Nat Commun* (2020) 11:4520. doi: 10.1038/s41467-020-18298-8

11. Byron A, Humphries JD, Humphries MJ. Defining the extracellular matrix using proteomics. *Int J Exp Pathol* (2013) 94:75–92. doi: 10.1111/iep.12011

12. Filipe EC, Chitty JL, Cox TR. Charting the unexplored extracellular matrix in cancer. *Int J Exp Pathol* (2018) 99:58–76. doi: 10.1111/iep.12269

13. Taha IN, Naba A. Exploring the extracellular matrix in health and disease using proteomics. *Essays Biochem* (2019) 63:417–32. doi: 10.1042/EBC20190001

14. Randles MJ, Humphries MJ, Lennon R. Proteomic definitions of basement membrane composition in health and disease. *Matrix Biol* (2017) 57-58:12–28. doi: 10.1016/j.matbio.2016.08.006

15. Dengjel J, Bruckner-Tuderman L, Nyström A. Skin proteomics - analysis of the extracellular matrix in health and disease. *Expert Rev Proteomics* (2020) 17:377–91. doi: 10.1080/14789450.2020.1773261

16. Rickelt S, Hynes RO. Antibodies and methods for immunohistochemistry of extracellular matrix proteins. *Matrix Biol* (2018) 71-72:10–27. doi: 10.1016/j.matbio.2018.04.011

17. Lindsey ML, Jung M, Hall ME, DeLeon-Pennell KY. Proteomic analysis of the cardiac extracellular matrix: clinical research applications. *Expert Rev Proteomics* (2018) 15:105–12. doi: 10.1080/14789450.2018.1421947

18. Kikuchi T, Hassanein M, Amann JM, Liu Q, Slebos RJ, Rahman SM, et al. In-depth proteomic analysis of nonsmall cell lung cancer to discover molecular targets and candidate biomarkers. *Mol Cell Proteomics* (2012) 11:916–32. doi: 10.1074/mcp.M111.015370

19. Zeng GQ, Zhang PF, Deng X, Yu FL, Li C, Xu Y, et al. Identification of candidate biomarkers for early detection of human lung squamous cell cancer by quantitative proteomics. *Mol Cell Proteomics* (2012) 11:M111.013946. doi: 10.1074/mcp.M111.013946

20. Nigro E, Imperlini E, Scudiero O, Monaco ML, Polito R, Mazzarella G, et al. Differentially expressed and activated proteins associated with non small cell lung cancer tissues. *Respir Res* (2015) 16:74. doi: 10.1186/s12931-015-0234-2

21. Hsu CH, Hsu CW, Hsueh C, Wang CL, Wu YC, Wu CC, et al. Identification and characterization of potential biomarkers by quantitative tissue proteomics of primary lung adenocarcinoma. *Mol Cell Proteomics* (2016) 15:2396–410. doi: 10.1074/mcp.M115.057026

22. Sandri BJ, Kaplan A, Hodgson SW, Peterson M, Avdulov S, Higgins L, et al. Multi-omic molecular profiling of lung cancer in COPD. *Eur Respir J* (2018) 52:1702665. doi: 10.1183/13993003.02665-2017

23. Zhou J, Liu B, Li Z, Li Y, Chen X, Ma Y, et al. Proteomic analyses identify differentially expressed proteins and pathways between low-risk and high-risk subtypes of early-stage lung adenocarcinoma and their prognostic impacts. *Mol Cell Proteomics* (2021) 20:100015. doi: 10.1074/mcp.RA120.002384

24. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest* (2017) 151:193–203. doi: 10.1016/j.chest.2016.10.010

25. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47:D330–8. doi: 10.1093/nar/gky1055

26. Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO. The extracellular matrix: tools and insights for the "omics" era. *Matrix Biol* (2016) 49:10–24. doi: 10.1016/j.matbio.2015.06.003

27. Shao X, Taha IN, Clauser KR, Gao Y, Naba A. MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Res* (2019) 48:D1136–44. doi: 10.1093/nar/gkz849

28. Tenzer S, Leidinger P, Backes C, Huwer H, Hildebrandt A, Lenhof HP, et al. Integrated quantitative proteomic and transcriptomic analysis of lung tumor and control tissue: a lung cancer showcase. *Oncotarget* (2016) 7:14857–70. doi: 10.18632/oncotarget.7562

29. Gocheva V, Naba A, Bhutkar A, Guardia T, Miller KM, Li CM, et al. Quantitative proteomics identify tenascin-c as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival. *Proc Natl Acad Sci USA* (2017) 114:E5625–34. doi: 10.1073/pnas.1707054114

30. Tian Y, Li H, Gao Y, Liu C, Qiu T, Wu H, et al. Quantitative proteomic characterization of lung tissue in idiopathic pulmonary fibrosis. *Clin Proteomics* (2019) 16:6. doi: 10.1186/s12014-019-9226-4

31. Bella J, Hulmes DJS. Fibrillar collagens. *Subcell. Biochem* (2017) 82:457–90. doi: 10.1007/978-3-319-49674-0_14

32. Schiller HB, Fernandez IE, Burgstaller G, Schaab C, Scheltema RA, Schwarzmayr T, et al. Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol Syst Biol* (2015) 11:819. doi: 10.15252/msb.20156123

33. Merl-Pham J, Basak T, Knüppel L, Ramanujam D, Athanason M, Behr J, et al. Quantitative proteomic profiling of extracellular matrix and site-specific collagen post-translational modifications in an *in vitro* model of lung fibrosis. *Matrix Biol Plus* (2019) 1:100005. doi: 10.1016/j.mbplus.2019.04.002

34. Wang Z, Zhai Z, Chen C, Tian X, Xing Z, Xing P, et al. Air pollution particles hijack peroxidasin to disrupt immunosurveillance and promote lung cancer. *eLife* (2022) 11:e75345. doi: 10.7554/eLife.75345

35. Kordowski F, Kolarova J, Schafmayer C, Buch S, Goldmann T, Marwitz S, et al. Aberrant DNA methylation of ADAMTS16 in colorectal and other epithelial cancers. *BMC Cancer* (2018) 18:796. doi: 10.1186/s12885-018-4701-2

36. Bartha Á., Győrffy B. TNMplot.com: a web tool for the comparison of gene expression in normal, tumor and metastatic tissues. *Int J Mol Sci* (2021) 22:2622. doi: 10.3390/ijms22052622

37. Jones MG, Andriotis OG, Roberts JJ, Lunn K, Tear VJ, Cao L, et al. Nanoscale dysregulation of collagen structure-function disrupts mechano-homeostasis and mediates pulmonary fibrosis. *eLife* (2018) 7:e36354. doi: 10.7554/eLife.36354

38. Brereton CJ, Yao L, Davies ER, Zhou Y, Vukmirovic M, Bell JA, et al. Pseudohypoxic HIF pathway activation dysregulates collagen structure-function in human lung fibrosis. *eLife* (2022) 11:e69348. doi: 10.7554/eLife.69348

39. Chen Y, Terajima M, Yang Y, Sun L, Ahn YH, Pankova D, et al. Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. *J Clin Invest* (2015) 125:1147–62. doi: 10.1172/JCI74725

40. Chen Y, Guo H, Terajima M, Banerjee P, Liu X, Yu J, et al. Lysyl hydroxylase 2 is secreted by tumor cells and can modify collagen in the extracellular space. *J Biol Chem* (2016) 291:25799–808. doi: 10.1074/jbc.M116.759803

41. Lu P, Takai K, Weaver VM, Werb Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb. Perspect Biol* (2011) 3:a005058. doi: 10.1101/cshperspect.a005058

42. Vizovišek M, Fonović M, Turk B. Cysteine cathepsins in extracellular matrix remodeling: extracellular matrix degradation and beyond. *Matrix Biol* (2019) 75-76:141–59. doi: 10.1016/j.matbio.2018.01.024

43. Burgess JK, Mauad T, Tjin G, Karlsson JC, Gunilla Westergren-Thorsson G. The extracellular matrix - the under-recognized element in lung disease? *J Pathol* (2016) 240:397–409. doi: 10.1002/path.4808

44. Passlick B, Sienel W, Seen-Hibler R, Wöckel W, Thetter O, Mutschler W, et al. Overexpression of matrix metalloproteinase 2 predicts unfavorable outcome in early-stage non-small cell lung cancer. *Clin Cancer Res* (2000) 6:3944–8.

45. Wang YZ, Wu KP, Wu AB, Yang ZC, Li JM, Mo YL, et al. MMP-14 overexpression correlates with poor prognosis in non-small cell lung cancer. *Tumour Biol* (2014) 35:9815–21. doi: 10.1007/s13277-014-2237-x

46. Hofmann HS, Hansen G, Richter G, Taege C, Simm A, Silber RE, et al. Matrix metalloproteinase-12 expression correlates with local recurrence and metastatic disease in non-small cell lung cancer patients. *Clin Cancer Res* (2005) 11:1086–92. doi: 10.1016/j.jtcvs.2017.11.110

47. Qu P, Du H, Wang X, Yan C. Matrix metalloproteinase 12 overexpression in lung epithelial cells plays a key role in emphysema to lung bronchioalveolar adenocarcinoma transition. *Cancer Res* (2009) 69:7252–61. doi: 10.1158/0008-5472.CAN-09-0577

48. Bhave G, Cummings CF, Vanacore RM, Kumagai-Cresse C, Ero-Tolliver IA, Rafi M, et al. Peroxidasin forms sulfilimine chemical bonds using hypohalous acids in tissue genesis. *Nat Chem Biol* (2012) 8:784–90. doi: 10.1038/nchembio.1038

49. Lázár E, Péterfi Z, Sirokmány G, Kovács HA, Klement E, Medzihradszky KF, et al. Structure-function analysis of peroxidasin provides insight into the mechanism of collagen IV crosslinking. *Free Radic Biol Med* (2015) 83:273–82. doi: 10.1016/j.freeradbiomed.2015.02.015

50. Schnellmann R, Sack R, Hess D, Annis DS, Mosher DF, Apte SS, et al. A selective extracellular matrix proteomics approach identifies fibronectin proteolysis by a disintegrin-like and metalloprotease domain with thrombospondin type 1 motifs (ADAMTS16) and its impact on spheroid morphogenesis. *Mol Cell Proteomics* (2018) 17:1410–25. doi: 10.1074/mcp.RA118.000676

51. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* (2012) 12:323–34. doi: 10.1038/nrc3261

52. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature* (2013) 501:355–64. doi: 10.1038/nature12627

53. Nieuwenhuis TO, Rosenberg AZ, McCall MN, Halushka MK. Tissue, age, sex, and disease patterns of matrisome expression in GTEx transcriptome data. *Sci Rep* (2021) 11:21549. doi: 10.1038/s41598-021-00943-x

54. Lennon R, Byron A, Humphries JD, Randles MJ, Carisey A, Murphy S, et al. Global analysis reveals the complexity of the human glomerular extracellular matrix. *J Am Soc Nephrol.* (2014) 25:939–51. doi: 10.1681/ASN.2013030233

55. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* (2003) 75:663–70. doi: 10.1021/ac026117i

56. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* (2016) 11:2301–19. doi: 10.1038/nprot.2016.136

57. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* (2014) 13:2513–26. doi: 10.1074/mcp.M113.031591

58. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* (2019) 47:D506–15. doi: 10.1093/nar/gky1049

59. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* (2016) 13:731–40. doi: 10.1038/nmeth.3901

60. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021Update). *Nucleic Acids Res* (2022) 50:W216–21. doi: 10.1093/nar/gkac194

61. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics* (2004) 20:1453–4. doi: 10.1093/bioinformatics/bth078

62. Saldanha AJ. Java Treeview–extensible visualization of microarray data. *Bioinformatics* (2004) 20:3246–8. doi: 10.1093/bioinformatics/bth349

63. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* (2010) 26:2927–8. doi: 10.1093/bioinformatics/btq562

64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13:2498–504. doi: 10.1101/gr.1239303

65. Heer J, Card SK, Landay JA. Prefuse: a toolkit for interactive information visualization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 421–30. New York, NY, USA: ACM. (2005) p. 421-30.

66. Lánczky A, Győrffy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *J Med Internet Res* (2021) 23: e27633. doi: 10.2196/27633